Instantaneous frequency estimation based on directionally smoothed pseudo-Wigner-Ville distribution bank

P.-L. Shui, H.-Y. Shang and Y.-B. Zhao

Abstract: In the cases of low signal-to-noise ratios (SNRs), estimating the instantaneous frequency (IF) curves of signals of interest is an interesting topic with many practical applications. Most of the existing methods are based on quadratic time-frequency (TF) distributions, which, however, yield a number of outliers in the cases of low SNRs. In this paper, we construct a new family of TF distributions, namely, the joint distributions, to estimate the IF curves in order to reduce outliers in the cases of low SNRs. The construction of the joint distributions is based on the definition of the directionally smoothed pseudo-Wigner-Ville distributions (DSPWVD) and pointwise adaptive weight averaging of a bank of DSPWVDs with different directions. The segments of the IF curve whose directions are close to that of the DSPWVD can be highlighted by each DSPWVD and the entire IF curve will be enhanced by the joint TF distribution. Simulation results show that the IF estimator based on the joint distributions outperforms that using quadratic TF distributions and the adaptive optimal kernel distribution in the cases of low SNRs.

1 Introduction

It is known that estimating the instantaneous frequency (IF) curves of unknown nonlinear frequency modulated (FM) signals in noise environments is an important issue with many practical applications; for example, in cases for which the radial velocities (instantaneous Doppler frequency shifts) of a manoeuvring target in radars need to be estimated from noisy echoes. There are a bundle of methods [1–4] based on time-frequency (TF) distributions available to estimate the IF curves. Among the methods, a commonly used approach is to detect the maxima positions of all time-slices of the TF distribution of a noisy signal (also called the TF energy ridge detection) [1–4], and the estimation performance is closely related to the underlying TF distributions and signal-to-noise ratios (SNRs).

It is well known that most TF distributions can highlight the TF energy ridges of signals from noise when SNRs are high, and in the cases of high SNRs, the estimation error is mainly composed of small bias resulted from the TF distribution itself and small random deviations inside the auto-term (AT) region from noise. The bias and random deviations can be reduced by using adequate TF distributions [3–5]. However, in the cases of low SNRs, the situation is quite different [6, 7]. In those cases, the TF energy ridge of a signal is often indistinct in the TF distributions due to strong noise and a few number of the values of the TF distributions outside the AT region probably surpass the values inside the AT region. Therefore, the IF curves estimated contain TF points outside the AT region. These points significantly deviate from the true IF curve and are referred to as outliers, which may be the primary factor [2] of performance degradation in the cases of low SNRs. It has been shown that the probability of the occurrence of the outliers depends on the maximal AT magnitudes and the distribution variance of a quadratic TF distribution [1], and the IF estimator using a quadratic TF distribution with a fixed kernel yields a number of outliers when SNR is lower than 0 dB [2], which is due to the fact that the two conflicting requirements, that is the small distribution variance and the large AT magnitudes for a nonlinear FM signal cannot be traded off by a fixed kernel.

The outliers can be reduced by using the TF distributions [8–14] that achieve a trade-off between the above-mentioned two requirements, where adaptive directional kernels are used to match the entire TF features or the local TF features of a signal. The adaptive directional kernels work in both the time-invariant and time-varying manners. The TF distributions with time-invariant adaptive kernels [9, 11–14] are suitable for the signals whose energy is distributed around a fixed TF direction, while the ones with time-varying adaptive kernels, or called the adaptive optimal kernel (AOK) distributions [8], provide better TF concentration and smaller distribution variance for the signals whose TF directions vary in a wide range with time. Alternatively, the linear TF transforms also work in the cases of low SNRs, which exploit atoms of TF directivity to compactly represent signals and extract the signal parameters [12, 13, 15–19]. This implies that it is an important way to exploit TF directivity in designing TF distributions so as to reduce the outlier of IF estimation in the cases of low SNRs.

In this paper, we define the directionally smoothed pseudo-Wigner-Ville distributions (DSPWVD), a smooth version of the pseudo-Wigner-Ville distribution (PWVD) along a special TF direction. From a bank of DSPWVDs with different directions, we construct a new family of TF distributions, namely, the joint distributions, by pointwise
adaptive weighted combination, which take the advantages of all the DSPWVDs. The merits of the joint distributions are using the adaptive kernels varying with TF points to concentrate the energy of a nonlinear FM signal whose TF directions vary in a wide range on its IF curve, in order that the outliers of the IF estimators using the joint distributions are reduced in the cases of low SNRs.

2 Directionally smoothed pseudo-Wigner-Ville distribution

Let \( x(n) \) be a noisy signal

\[
  x(n) = s(n) + w(n) = a \exp(2\pi j\phi(n)) + w(n) \tag{1}
\]

where \( s(n) \) is the signal of interest and \( w(n) \) is zero-mean i.i.d white complex Gaussian noise of variance \( \sigma^2 \). The IF curve of the signal can be estimated by using the maxima position detection to the noisy TF distribution (also called the TF energy ridge detection), that is

\[
  IF(n) = \arg \max_{f \in [-1/4, 1/4]} \{ C_x(n, f; \varphi) \},
\]

\[
  C_x(n, f; \varphi) = \sum_{m, k} \varphi(m, k) x(n + m + k) x^*(n + m - k) \exp(-4\pi j k)
\]

where \( C_x(n, f; \varphi) \) is a discrete-time TF distribution of Cohen class with a time-lag kernel \( \varphi(m, k) \).

The estimation error of (2) is composed of bias, small random deviations inside the AT region of the signal, and large random deviations outside the AT region (also called outliers in an estimate). In the cases of high SNRs, large random deviations infrequently occur and most TF distributions can be used to obtain satisfactory results. In the cases of low SNRs, the outliers are the primary factor of performance degradation. Up to now, only a few kinds of TF distributions can work well in the cases of low SNRs, to give an example, the AOK distributions [8]. The probability of the occurrence of the outliers depends on the ratio of the maximal AT magnitudes of the signal to the distribution variance. The more this ratio is, the less the probability of the occurrence of the outliers is. For the noisy signal in (1), the mean and the distribution variance of a TF distribution of Cohen class are [20]

\[
  E\{ C_x(n, f; \varphi) \} = C_x(n, f; \varphi) + \sigma^2 \sum_m \varphi(m, 0),
\]

\[
  \text{Var}\{ C_x(n, f; \varphi) \} = (2\sigma^2 + \sigma^4) \sigma^2 \sum_k |\varphi(m, k)|^2
\]

(3)

In the cases of low SNRs, the distribution variance depends on the kernel, whereas the maximal AT magnitudes of the signal depend on the TF concentration of the TF distribution to the signal. The second-order derivatives of the phase function \( \phi(n) \) of the signal are referred to as the chirp rates and the chirp rates geometrically correspond to the slopes of the tangential lines of the IF curve at an instant.

When the chirp rates of a signal vary in a wide range, the TF direction of a fixed kernel cannot match the chirp rates of a signal at all IF points. Consequently, a TF distribution of Cohen class cannot achieve large values at all IF points and the corresponding IF estimator yields a number of outliers.

Practically, we always assume that the IF curve of a nonlinear FM signal is a smooth curve that can be approximated by a piecewise linear function. This implies that the signal segments in short-time intervals are approximate linear

FM (LFM). Thus, a PWVD with an adequate lag window can coherently integrate the signal energy in the window at the IF point and the value of the PWVD at the IF point is large. This can be interpreted in detail using an example. Let the signal in a sliding window \([-K + n, K + n]\) satisfy \( s(n + k) \approx \exp(2\pi j(\eta_0 + \eta_1 k + 1/2\pi k^2)) \), for \( k = -K, \ldots, K \), where \( 2\pi \eta_0 \), \( \eta_1 \) are the initial phase, the instantaneous frequency, and the chirp rate of the signal at the instant \( n \), respectively. Then, at the IF point \((n, \eta_1)\), the value of the PWVD with a rectangle window of length \( 2K + 1 \)

\[
  \text{PWVD}_x(n, \eta_1) = \sum_{k=-K}^{K} s(n + k)s^*(n - k) \exp(-4\pi jk \eta_1) \approx (2K + 1)a^2
\]

(4)

This indicates that the PWVD realizes locally coherent integration of the signal energy. However, the noisy PWVD has a large distribution variance, \((2K + 1)a^2(2a^2 + \sigma^2)\), which has detrimental effect on the IF estimation in the cases of low SNRs. Smoothing the PWVD along the time axis derives a smoothed PWVD (SPWVD) with a smaller distribution variance. The SPWVD with a rectangle smooth window of length \( 2M + 1 \), \( a(m) = 1/(2M + 1) \), \( m = 0, \pm 1, \ldots, M \), has a smaller distribution variance, \((2K + 1)a^2(2a^2 + \sigma^2)/(2M + 1)\). Note that smoothing the PWVD along the time axis does not incur significant reduction of the values at the IF points, only when the chirp rates of a signal segment are close to zero. This implies that the SPWVD is only suited to detect the segments approximately parallel to the time axis on the IF curve.

2.1 Directionally smoothed pseudo-Wigner-Ville distribution

A general form of the SPWVDs is

\[
  \text{SPWVD}_x(n, f; h, g) = \sum_m g(m) \text{PWVD}_x(n + m, f, h) = \sum_m g(m) \sum_k h(k) x(n + m + k) x^* (n + m - k) \exp(-4\pi j k f) \tag{5}
\]

where and \( g \) and \( h \) are the lag window and time smooth window, respectively. A SPWVD is obtained by smoothing the PWVD along the time axis. In order to highlight the segments along the direction \( \theta \) on the IF curve in the cases of low SNRs, we smooth the PWVD along the direction \( \theta \), which derives a DSPWVD. A DSPWVD is defined by

\[
  \text{DSPWVD}_x(n, f; h, g) = \sum_m g(m) \text{PWVD}_x(n + m, f + m \tan \theta; h) = \sum_m g(m) h(k) x(n + m + k) x^* (n + m - k) \exp(-4\pi j k f) \tag{6}
\]

It belongs to the TF distributions of Cohen class and has a time-lag kernel specified by the directional angle \( \theta \), the lag window \( h \) and the smooth window \( g \)

\[
  \varphi(m, k; \theta) = g(m) h(k) \exp(-4\pi j k m \tan \theta) \tag{7}
\]
According to (3), the mean and the distribution variance of DSPWVD$^\delta (n, f; g, h)$ of a noisy signal $x(n)$ in (1) are

$$
E\{\text{DSPWVD}^\delta (n, f; g, h)\} = \text{DSPWVD}^\delta (n, f; h, g) + a^2 h(0) \sum_m g(m),
$$

$$
\text{Var}\{\text{DSPWVD}^\delta (n, f; g, h)\} = \left(2a^2 + a^2\right) \varphi^2 \sum_m g^2(m) \sum_k h^2(k) \tag{8}
$$

Note that the distribution variance is independent of $\theta$ and the mean is the sum of the two terms, where the first term, the DSPWVD of the signal, is sensitive to the directional angle $\theta$, and the second term is independent of $\theta$ and the signal.

### 2.2 Optimal window design in DSPWVD

Following (8), in a noisy DSPWVD, the TF energy ridge of a signal is superimposed on a randomly fluctuating pedestal with an average height $a^2 h(0) \sum_m g(m)$ and a fluctuating variance $\left(2a^2 + a^2\right) \varphi^2 \sum_m g^2(m) \sum_k h^2(k)$. In order to easily detect the IF curve from the noisy DSPWVD, it is desired to have both large values on the IF curve and a small distribution variance. The distribution variance is dependent on the windows and SNR and is independent of the signal. The values of the DSPWVD of the signal on the IF curve are closely related to the smooth direction and the chirp rates of the signal. Let $x(n) = axe^{-2\pi j (n - \eta_1 + \eta_2 n)}$ be a LFM signal with a chirp rate $\eta_2$, the value of the DSPWVD at the IF point $(n, \eta_1 + \eta_2 n)$ is

$$
\text{DSPWVD}^\delta (n, \eta_1 + \eta_2 n) = a^2 \sum_m \sum_k h(k) \exp (-4\pi j km (\tan \theta - \eta_2) g(m)) \tag{9}
$$

When the lag window $h(k)$ is symmetric with respect to $k = 0$, PWVD, and DSPWVD$^\delta$ are real for any direction $\theta$, and (9) is simplified to

$$
\text{DSPWVD}^\delta (n, \eta_1 + \eta_2 n) = a^2 \sum_m \sum_k h(k) \cos (4\pi k m (\eta_2 - \tan \theta)) g(m) \tag{9'}
$$

This implies that the value of the DSPWVD at the IF point is large when $|\eta_2 - \tan \theta|$ is small while the value is small when $|\eta_2 - \tan \theta|$ is large. Note that (9') is also valid when the signal segment in a sliding window is approximate LFM.

The above analysis shows that the DSPWVD can highlight the IF points whose chirp rates are close to $\tan \theta$. In other words, it is suited to detect the IF points whose chirp rates lie in an interval $[\tan \theta - \Delta, \tan \theta + \Delta]$, where $\Delta$ is a small positive number. In order to minimise the probability of the occurrence of the outliers in detecting the IF points whose chirp rates lie in $[\tan \theta - \Delta, \tan \theta + \Delta]$, the optimal lag window and smooth window can be designed by maximising the ratio of the average magnitude of the DSPWVD of the signal at these IF points to the distribution standard deviation. For convenience, assume the chirp rates to be uniformly distributed in the interval $[\tan \theta - \Delta, \tan \theta + \Delta]$. In this way, the average magnitude of the DSPWVD of the signal at these TF points is formulised as

$$
\text{AM}(h, g, \Delta) = \frac{1}{2\Delta} \int_{\tan \theta - \Delta}^{\tan \theta + \Delta} \text{DSPWVD}^\delta (n, \eta_1 + \eta_2 n) d\eta_2
$$

$$
= a^2 \sum_m \sum_k h(k) g(m) \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} \cos (4\pi k m \Delta) dz \tag{10}
$$

where $\text{sinc}(x) = \sin(\pi x) / (\pi x)$. The ratio of the average magnitude to the distribution standard deviation is formulised as

$$
\frac{\text{AM}(h, g, \Delta)}{\sqrt{(2a^2 + a^2) \varphi^2 \|g\|_2^2 \|h\|_2^2}} = \frac{\sum_m \sum_k h(k) \text{sinc}(4\pi k m \Delta) g(m)}{\sqrt{(2a^2 + a^2) \varphi^2} \text{Gain}(h, g, \Delta)}
$$

The first term in the above formula is determined by the SNR and the second term is determined by the two windows and $\Delta$. We refer the second term as the integration gain of the DSPWVD for IF detection. The optimal windows are designed by maximising the integration gain, that is,

$$
\max_{h,g} \text{Gain}(h, g, \Delta) = \left\{ \begin{array}{l}
\frac{h^T A \Delta g}{\|h\|_2 \|g\|_2} \\
\end{array} \right. \tag{11}
$$

where the lag window is assumed to support in $\{-K, -K+1, \ldots, K\}$, the smooth window in $\{-M, -M+1, \ldots, M\}$, and

$$
A_\Delta = [\text{sinc}(4\pi m \Delta)]_{m=-M,-M+1,\ldots,M}^{m=M-M+1,\ldots,M} \tag{12}
$$

is a $(2K + 1)$ by $(2M + 1)$ matrix.

In terms of the singular value decomposition (SVD) theory [21]

$$
\max_{h,g} \text{Gain}(h, g, \Delta) = \sigma_{\text{max}}(A_\Delta)
$$

where $\sigma_{\text{max}}(A_\Delta)$ is the largest singular value of $A_\Delta$. The optimal lag window and the optimal smooth window are the left singular vector and the right singular vector associated with the largest singular value, respectively. The structure of $A_\Delta$ also assures symmetry of the optimal windows.

For convenience, we constraint $h(0) = 1$ and $\sum_m g(m) = 1$.

For $K = 64, M = 32$ and $\Delta$ varies from 0.0001 to 0.001, the integration gains of the optimal windows and several commonly used windows are plotted in Fig. 1(a). Fig. 1(b) illustrates the optimal lag window when $\Delta = 0.0001$, the rectangle, Hamming, Gauss and Kaiser windows. Except the optimal windows, the same shaped smooth windows of length 65 as the lag windows are used to calculate the integration gains. It can be seen that the optimal windows provide the largest integration gains. Particularly, the integration gains of the rectangle windows are close to the largest integration gains when $\Delta$ is very small. As a result, the optimal windows can be replaced by the rectangle windows when $\Delta$ is very small, in order to simplify the calculation of DSPWVDs. We have also seen that the optimal windows are independent
of direction $\theta$, in other words, the optimal windows are valid for different smooth directions, which enable us to simplify the computation of the sequent joint distributions.

3 Joint distribution and if curve estimation

The SPWVD can also be interpreted as smoothing the discrete Wigner-Ville distribution using a two-dimensional (2-D) directional filter in the TF plane. Set

$$g_{m}(f) = \frac{1}{C_{0}} \exp(-4\pi k(f - m \tan \theta))$$

be a 2-D directional filter in the TF plane specified by the windows $g(m)$ and $h(k)$, and the angle $\theta$. It is easily proved from the definition (6) that

$$\text{DSPWVD}_{\theta}(n, f) = \text{WVD}_{\theta}(n, f) \otimes \Pi_{g,h}(n, f)$$

(13)

be a joint distribution, where $\text{WVD}_{\theta}(n, f) = \sum_{k} h(k)w(n + k)w^*(n - k)e^{-4\pi kf}$ and the symbol $\otimes$ denotes the 2-D convolution that is discrete in time and continuous in frequency.

DSPWVDs exhibit different TF directivity with variant of the angle $\theta$. For the optimal windows with $K = 64$, $M = 32$ and $\Delta = 0.0001$, the directional filters with $\tan \theta = 0.001, 0.0005, 0$ and $-0.001$ are illustrated in Fig. 2. Each DSPWVD can highlight the IF points whose chirp rates close to $\tan \theta$ from the fluctuant noise pedestal, which is owing to that the corresponding directional filter matches the chirp rates of these IF points. Estimating the IF curve of a signal in noise, a TF distribution is desired to highlight all IF points from the fluctuant noise pedestal. In what follows, we combine a bank of DSPWVDs with different smooth directions in order to generate a new family of TF distributions that can highlight the entire IF curve of a nonlinear FM signal whose chirp rates vary in a wide range.

3.1 Joint distributions and efficient computation

Assume that the chirp rates of a nonlinear FM signal vary in the interval $(\tan \alpha_1, \tan \alpha_2)$. We uniformly segment the interval into $L + 1$ subintervals

$$[\tan \alpha_1 + l\Delta, \tan \alpha_1 + (l + 1)\Delta], \quad l = 0, 1, \ldots, L,$$

$$\Delta = (\tan \alpha_2 - \tan \alpha_1)/L$$

Take a set of direction angles

$$\theta_l = \arctan(\tan \alpha_1 + (l + 0.5)\Delta), \quad l = 0, 1, \ldots, L \quad (15)$$

Herein, the directions are discretised by uniformly sampling $\tan \theta$ rather than angle $\theta$, because the integration gain of a DSPWVD at the IF points is a function of the distance $|\tan \theta - \eta|$.

In what follows, we deal with the combination problem of a bank of DSPWVDs with different smooth directions. It is known that the optimal combination of multiple spectrograms with different windows can overcome some...
limitations of a single spectrogram [22, 23]. Here, we construct the joint distribution from DSPWVD\(\beta(n, f)\), \(l = 0, 1, \ldots, L\) by pointwise adaptive weighted combination. The joint distribution with parameter \(\beta\) is defined by
\[
JD^\beta(n, f) = \sum_{l=0}^{L} W_\beta(n, f, l)DSPWVD^\beta(n, f), \quad 0 \leq \beta < +\infty
\]
(16)
where every TF point is arranged a weight
\[
W_\beta(n, f, l) = \frac{(\max\{DSPWVD^\beta(n, f), 0\})^\beta}{\sum_{l=0}^{L} (\max\{DSPWVD^\beta(n, f), 0\})^\beta}.
\]
(17)
Particularly, when \(\beta \to +\infty\), we have
\[
\lim_{\beta \to +\infty} JD^\beta(n, f) = JD^\infty(n, f)
\]
\[
= \max\{DSPWVD^\infty(n, f)\} = \max\{x(n)\}
\]
(18)
For every TF point, JD\(\infty(n, f)\) is the largest order statistics of the \((L + 1)\) random variables DSPWVD\(\infty(n, f)\). It is easily observed that the joint distribution uses adaptive kernels varying with TF points and the adaptive kernel at each TF point is
\[
\varphi_\beta(n, k) = \sum_{l=0}^{L} W_\beta(n, k, l)\varphi(n, k; \theta_l)
\]
(19)
As a result, a joint distribution preserves large values at all IF points of a signal because the adaptive kernels varying with TF points can better track the local TF directions of the signal even in the cases of low SNRs. Another problem relevant to the capability of noise suppression is whether a joint distribution has a small distribution variance. Outside the AT region, JD\(\beta(n, f)\) is the weighed sum (\(\beta \neq \infty\)) or the largest order statistics (\(\beta = \infty\)) of the \((L + 1)\) random variables of mean \(\sigma^2\) and variance \(\sigma^2(2a^2 + \sigma^2)|h||g|^2\). These \((L + 1)\) random variables have too complicated dependency to give an analytic form of the probability density function (PDF) of the values of the joint distribution outside the AT region [24]. The simulations show that the joint distributions have smaller distribution variances. Consequently, the joint distributions have large values on the IF curve and smaller distribution variances, which reduce the probability of the occurrence of the outliers in the cases of low SNRs.

In order to demonstrate the detection capability of the joint distribution, we compare it with several commonly used TF distributions for a cosine FM signal corrupted by additive white complex Gaussian noise (AWCGN)
\[
x(n) = as(n) + w(n)
\]
\[
= a\exp(14\pi\sin(0.012n)) + w(n)
\]
(20)
where \(s(n)\) is the signal of interest, \(w(n)\) is a zero-mean AWCGN of unit variance. In the joint distribution, \([\tan\alpha_1, \tan\alpha_2] = [-0.001, 0.001]\), \(\Delta = 0.0001\), \(\beta = +\infty\), and the optimal windows are used. When SNR = \(-7.5\) dB, the PWVD, SPWVD, the AOK distribution and the joint distribution are plotted in Fig. 3. Obviously, it is difficult to identify the IF curve in the PWVD and SPWVD because the PWVD has a large distribution variance and the SPWVD does not preserve large values at all IF points. Owing to using time-varying adaptive kernels and adaptive kernels varying with the TF points, the IF curve is readily identified in the AOK distribution and the joint distribution. More accurately, the detection capability of a TF distribution can be evaluated using the PDFs of its values on the IF curve and outside the AT region [1]. The lesser the two PDFs of a distribution are overlapped, the

![Fig. 3](image-url)  
**Fig. 3** PWVD, SPWVD, the AOK distribution, and the joint distribution of a noisy signal

\(a\) PWVD of the noisy signal (SNR = \(-7.5\) dB)

\(b\) SPWVD of the optimal windows of length 129 and 65

\(c\) AOK distribution with sliding analysis window of length 128 [8]

\(d\) Joint distribution \(JD^\beta\) of \(K = 63, M = 32, L = 20\) and \(\Delta = 0.0001\)
losser is the probability of the occurrence of the outliers in the corresponding estimator. Here, for the test signal in (20) and computer-generated AWGN, 200 independent Monte-Carlo tests used to estimate the two empirical PDFs. As shown in Fig. 4, the two empirical PDFs of the PWVD and the SPWVD are overlapped badly, meaning that the corresponding IF estimators suffer from significant probabilities of the occurrence of the outliers. The two empirical PDFs of the joint distribution are overlapped less than that of the AOK distribution do, indicating that the IF estimator using the joint distribution should have a lesser probability of the occurrence of the outliers than one using the AOK distribution does.

The computational complexity needs to be considered with practical applications. The computational cost of an IF estimator using a TF distribution mainly comes from the calculation of the TF distribution itself. The computational complexity of a quadratic TF distribution of a joint distribution is \( O(NP^2) \) times complex multiplications. The joint distribution requires to compute \( (L + 1) \) DSPWVDs and \( (L + 1) \) weights at each TF point. If all the DSPWVDs are calculated separately, then \( O((L + 1)NP) \) times complex multiplications are required. Computation of all weights requires \( (L + 1)NP \) times real divisions and multiplications. As a result, the computational complexity of a joint distribution is \( O((L + 1)NP\log P) \), a heavy computational burden.

In what follows, we give an efficient algorithm to compute the joint distribution. It is notable that all DSPWVDs use the same lag window and the smooth window, which can be exploited to reduce the computational cost. The proposed algorithm is realised in three steps. In the first step, a discrete-frequency PWVD with the lag window \( h(n) \) is calculated and stored as a \( N \) by \( P \) matrix

\[
X(n, p) = \text{PWVD}(n, p\Delta f - 1/4; h),
\]

where \( \Delta f \) is the frequency sampling interval. In the second step, calculate

\[
Y_i(n, p) = \text{DSPWVD}(n, p\Delta f - 1/4; h, g)
\]

\[
= \sum_m g(m)\text{PWVD}(n + m, p\Delta f + m\tan\theta_i - 1/4; h)
\]

where \( \theta_i = \text{round}(m\tan\theta_i/\Delta f) \) is the integer nearest to \( m\tan\theta_i/\Delta f \). The directional smoothing in (6) and (14) is replaced by smoothing \( X(n, p) \) along a discrete line segment in the discrete TF lattice. Due to symmetry of the smooth window, (23) is simplified to

\[
Y_i(n, p) \approx g(0)X(n, p) + \sum_{m=1}^{M} g(m)
\]

\[
\times [X(n - m, p - m) + X(n + m, p + m)]
\]

In this way, all the \( (L + 1) \) DSPWVDs are calculated from \( (M + 1) \) matrices \( g(m)X, m = 0, 1, \ldots, M \) by addition of their shift versions in \( (M + 1)NP \) times real multiplications. If using a rectangle smooth window, then the DSPWVDs can be calculated without multiplications. In the third step, the discrete joint distribution \( Z^\beta(n, p) \) is calculated from the matrices \( Y_i(n, p), l = 0, 1, \ldots, L \). In conclusion, the total computational cost is \( O(NP\log P) \) times complex multiplications for the PWVD, \( (L + 1)NP \) times real divisions for the weights and \( (M + L + 2)NP \) times real multiplications for \( (L + 1) \) DSPWVDs and the weighted summation.

### 3.2 IF curve estimation using joint distribution

Applying the maxima position detector to the joint distribution

\[
Z^\beta(n, p) \simeq JD^\beta(n, p\Delta f - 1/4), \quad n = 1, 2, \ldots, N,
\]

\[
p = 1, 2, \ldots, P
\]

The IF curve of a signal is estimated by

\[
\hat{\text{IF}}(n) = \Delta f \arg \max_p \{ Z^\beta(n, p) \} - 1/4
\]

It is known that the probability of the occurrence of the outliers depends on overlapping extent of the values of the distribution on the IF curve and outside the AT region. For the joint distributions, the overlapping extent is related to the parameter \( \beta \) and the TF feature of signals. When \( \beta \) is close to zero, the adaptive kernels degenerate to a fixed kernel, and the values of the joint distributions on the IF curve are small because the kernels cannot efficiently track the local TF direction of the signal. When \( \beta \) is large or positive infinite, the values of the joint distribution on the IF curve are large but the distribution variance also becomes large. As shown in Fig. 5, the empirical PDFs of the values of the four joint distributions with \( \beta = 0.5, 1.2, \infty \) outside the AT region and on the IF curve are plotted for the test signal (20) (SNR = -7.5 dB). Obviously, as increase of the parameter \( \beta \), the variances of the values outside the AT region become large while the means of the values on the IF curve also become large. The former variation is detrimental, while the latter variation is advantageous to IF estimation. Therefore, it is difficult to give a theoretic answer on which \( \beta \) is the best for the IF estimation. Fortunately, the sequent simulations show that the performance of the IF estimators is robust to \( \beta \) variation in a wide range.
has a polynomial phase function and its expression is cosine FM signal given in (20), and the second test signal of the occurrence of the outliers for each SNR level. In the simulations, 100 outliers, the mean squared error (MSE) is unsuited to evaluate points unreliable. Due to impulse characteristic of the outliers, the probability of the occurrence of the outliers is calculated by in the joint distribution. In (26), we exclude two boundaries because the boundary effect in the TF distribution is 0.5/(2K + 1) in the normalised frequency domain. The AT region is defined as

$$R_{AT} = \{(m, f) : |f - \text{IF}(n)| \leq 1/(2K + 1), |f| \leq 1/4\} \quad (25)$$

In the simulations below, the AOK distribution uses the sliding window of length 128 and the joint distribution uses the optimal lag window of length 127 ($K = 64$). The probability of the occurrence of the outliers is calculated by

$$P_f = (N - 2K)^{-1} \{ |\text{IF}(n) - \text{IF}(n)| > 1/128, \quad n = K + 1, K + 2, \ldots, N - K \} \quad (26)$$

where $\text{IF}(n)$ and $\overline{\text{IF}}(n)$ is the true and estimated IF curves, respectively, and $\Phi[A]$ denotes the number of samples satisfying the condition $A$, $K$ is the half of the lag window in the joint distribution. In (26), we exclude 2K points near the two boundaries because the boundary effect in the TF distribution (as shown in Fig. 3) makes these estimates of these points unreliable. Due to impulse characteristic of the outliers, the mean squared error (MSE) is unsuited to evaluate the performance of IF estimators. In the simulations, 100 times independent tests are made to calculate the probabilities of the occurrence of the outliers for each SNR level.

In the simulation, we use two test signals, the first is the cosine FM signal given in (20), and the second test signal has a polynomial phase function and its expression is

$$s_t(n) = a \exp(2 \pi ( -20(n/512 - 1) - 25(n/512 - 1)^2 
+ 20(n/512 - 1)^3 + 20(n/512 - 1)^4 )) \quad (27)$$

where $n = 1, 2, \ldots, 1024$. The chirp rates of the two signals lie in the interval $[-0.001, 0.001]$. In the AOK distribution, the length of sliding window is 128 and the normalised volume of the adaptive kernels is 3 [8]. The joint distributions of $\beta = 0.5, 1, 2, 4, 7, \infty$ are used, where $[\tan^{-1} \sigma_1, \tan^{-1} \sigma_2] = [-0.001, 0.001]$, $\Delta = 0.0001$ and $L = 20$. For the two test signals, the probabilities of the occurrence of the outliers are shown in Fig. 6 for SNRs from $-5$ to $-9$ dB. From Fig. 6, we draw conclusions as follow. First, the IF estimators using the joint distributions yield less outliers than one using the AOK distribution does in the cases of low SNRs. Secondly, the estimators using the joint distributions with $\beta \in [1, \infty]$ are robust in performance to variation of the parameter $\beta$. Based on the second conclusion, we prefer to use the $JD_2^*$ and $JD_\infty^*$ with practical applications because the power operations are avoided in calculating their adaptive weights. Additionally, we also found that the outliers unfrequently occur in all the six estimators, when SNRs are more than $-4$ dB.

The number $(L + 1)$ of directions in the joint distribution determines computational cost. A smaller number $(L + 1)$ results in a lower the computational cost. The smaller $(L + 1)$, however, also results in a coarser directional search, which will degrade the performance of the IF estimator. For example, when the joint distribution with $L = 10$ and $\beta = 1$ is applied to the first test signal, the probabilities of the occurrence of the outliers are 0.0041, 0.0174, 0.0541, 0.1342 and 0.2436 for SNRs $=-5, -6, -7, -8$ and $-9$ dB respectively; when the joint distribution with $L = 20$ and $\beta = 1$ is used, the probabilities are 0.0016, 0.064, 0.0286, 0.0719 and 0.1788 for SNRs $=-5, -6, -7, -8$ and $-9$ dB, respectively. This indicates that the number $(L + 1)$ of the directions in the joint distributions is substantially a trade-off between computational cost and performance.

In what follows, we analyse the reason that the joint distributions outdo the AOK distribution. In principle, the two kinds of distributions use the adaptive directional kernels to track the local TF features of a signal. However, they use the different tactics to find the adaptive directional kernels. The AOK distributions use the time-varying adaptive kernels that are determined from the short-time ambiguity functions (STAF) of a noisy signal segment within a sliding window. All noise within the sliding window disturbs the directivity of the kernels by its noisy

### 4 Experiment results

In the cases of low SNRs, the probability of the occurrence of the outliers is used to evaluate the performance of an IF estimator. For a nonlinear FM signal with an IF curve $\text{IF}(n)$, the AT region is defined as a strip region whose midline is the true IF curve. The width of the strip depends on the length of the lag window of the TF distribution. When the length of the lag window is $2K + 1$, the critical frequency sampling interval (the frequency resolution) of the PWVD is $0.5/(2K + 1)$ in the normalised frequency domain. The AT region is defined as

$$R_{AT} = \{(m, f) : |f - \text{IF}(n)| \leq 1/(2K + 1), |f| \leq 1/4\} \quad (25)$$

In the simulations below, the AOK distribution uses the sliding window of length 128 and the joint distribution uses the optimal lag window of length 127 ($K = 64$). The probability of the occurrence of the outliers is calculated by

$$P_f = (N - 2K)^{-1} \{ |\overline{\text{IF}}(n) - \text{IF}(n)| > 1/128, \quad n = K + 1, K + 2, \ldots, N - K \} \quad (26)$$

where $\text{IF}(n)$ and $\overline{\text{IF}}(n)$ is the true and estimated IF curves, respectively, and $\Phi[A]$ denotes the number of samples satisfying the condition $A$, $K$ is the half of the lag window in the joint distribution. In (26), we exclude 2K points near the two boundaries because the boundary effect in the TF distribution (as shown in Fig. 3) makes the estimates of these points unreliable. Due to impulse characteristic of the outliers, the mean squared error (MSE) is unsuited to evaluate the performance of IF estimators. In the simulations, 100 times independent tests are made to calculate the probabilities of the occurrence of the outliers for each SNR level.

In the simulation, we use two test signals, the first is the cosine FM signal given in (20), and the second test signal has a polynomial phase function and its expression is

$$s_t(n) = a \exp(2 \pi ( -20(n/512 - 1) - 25(n/512 - 1)^2 
+ 20(n/512 - 1)^3 + 20(n/512 - 1)^4 )) \quad (27)$$

where $n = 1, 2, \ldots, 1024$. The chirp rates of the two signals lie in the interval $[-0.001,0.001]$. In the AOK distribution, the length of sliding window is 128 and the normalised volume of the adaptive kernels is 3 [8]. The joint distributions of $\beta = 0.5, 1, 2, 4, 7, \infty$ are used, where $[\tan^{-1} \sigma_1, \tan^{-1} \sigma_2] = [-0.001, 0.001]$, $\Delta = 0.0001$ and $L = 20$. For the two test signals, the probabilities of the occurrence of the outliers are shown in Fig. 6 for SNRs from $-5$ to $-9$ dB. From Fig. 6, we draw conclusions as follow. First, the IF estimators using the joint distributions yield less outliers than one using the AOK distribution does in the cases of low SNRs. Secondly, the estimators using the joint distributions with $\beta \in [1, \infty]$ are robust in performance to variation of the parameter $\beta$. Based on the second conclusion, we prefer to use the $JD_2^*$ and $JD_\infty^*$ with practical applications because the power operations are avoided in calculating their adaptive weights. Additionally, we also found that the outliers unfrequently occur in all the six estimators, when SNRs are more than $-4$ dB.

The number $(L + 1)$ of directions in the joint distribution determines computational cost. A smaller number $(L + 1)$ results in a lower the computational cost. The smaller $(L + 1)$, however, also results in a coarser directional search, which will degrade the performance of the IF estimator. For example, when the joint distribution with $L = 10$ and $\beta = 1$ is applied to the first test signal, the probabilities of the occurrence of the outliers are 0.0041, 0.0174, 0.0541, 0.1342 and 0.2436 for SNRs $=-5, -6, -7, -8$ and $-9$ dB respectively; when the joint distribution with $L = 20$ and $\beta = 1$ is used, the probabilities are 0.0016, 0.064, 0.0286, 0.0719 and 0.1788 for SNRs $=-5, -6, -7, -8$ and $-9$ dB, respectively. This indicates that the number $(L + 1)$ of the directions in the joint distributions is substantially a trade-off between computational cost and performance.

In what follows, we analyse the reason that the joint distributions outdo the AOK distribution. In principle, the two kinds of distributions use the adaptive directional kernels to track the local TF features of a signal. However, they use the different tactics to find the adaptive directional kernels. The AOK distributions use the time-varying adaptive kernels that are determined from the short-time ambiguity functions (STAF) of a noisy signal segment within a sliding window. All noise within the sliding window disturbs the directivity of the kernels by its noisy
STAF, which results in that the kernel’s direction cannot better track the local TF direction of a signal in the cases of low SNRs. The joint distributions use the adaptive kernels varying with TF points that are obtained by adaptive weighted combination of \((L + 1)\) predefined directional kernels. Only a narrow-band noise whose energy is distributed near the TF point disturbs the directivity of the kernels. Thus, the kernel’s directions in the joint distributions can better track the local TF directions of a signal in the cases of low SNRs.

5 Conclusion

In the paper, we constructed a family of joint distributions from a DSPWVD bank with different directions. The joint distributions use the adaptive kernels varying with TF points that can efficiently track the local TF directions of a nonlinear FM signal in the cases of low SNRs. The joint distributions have large values on the IF curve and small distribution variances. The IF estimators using the joint distributions have less probabilities of the occurrence of the outliers than ones using other distributions do, in the cases of low SNRs.

In addition, the adaptive kernels varying with TF points in the joint distributions should be advantageous to detection of the IF curves of multicomponent signals. For example, when two components have different instantaneous frequencies and different chirp rates at an instant, the joint distributions can use the two different shaped kernels to match the chirp rates of the two components at the instant, whereas the AOK distribution cannot but use one kernel to the two components. In this way, the joint distribution has larger values at the two IF points than the AOK distribution does, which will help to reduce the occurrence of outliers in estimating the IF curves of multicomponent signals.

The maxima position detection of the TF distribution is a simple but also efficient method to estimate IF curve of a signal in noise environment. In this method, all IF points are independently detected from the corresponding time-slices of the TF distribution, which implies that the maxima position detection does not exploit the dependency in locations among adjacent IF points. In fact, since the IF curve of a nonlinear FM signal is a smooth curve, the IF points at adjacent instants are near each other. In [25], the hidden Markov model (HMM) is used to model the transition rule of the adjacent IF points and the corresponding frequency track algorithm based on the HMM can reject highly improbable outliers. Intuitively, it is possible to further reduce outliers by means of combining the maxima position detection of the joint distribution with the HMM described the transition rule of the IF points at the adjacent instants.

6 References